COLORADO STATE
UNIVERSITY
FORT COLLINS, COLORADO
80523

# department of electrical engineering

14

AD A024279

(12)

STOCHASTIC APPROXIMATION WITH CORRELATED DATA

Technical Report

by

David C. Farden

D D C
MAY 10 1976

| | |
|---|---|
| ARPA Order Number: | 1872 |
| Contract Number: | 10292750 |
| Program Code No.: | 5G10 |
| Naval Undersea Center Contract Number: | N66001-75-C-0224 |
| Principal Investigators: | David C. Farden (303) 491-6018 |
| | Louis L. Scharf (303) 491-5065 |
| Name of Contractor: | Colorado State University |
| Project Engineer: | Darrell E. Marsh (714) 225-6135 |
| Effective Date of Contract: | 30 May 1975 |
| Short Title of Work: | Communication Receiver Studies-II |
| Contract Expiration Date: | 1 March 1976 |
| Amount of Contract: | $14,998.00 |

Sponsored in part by
Advanced Research Projects Agency
ARPA Order No. 1872

ACKNOWLEDGMENT

This research was supported in part by the Advanced Research Projects Agency of the Department of Defense and monitored by the Naval Undersea Center, San Diego, under Contract Numbers N66001-74-C-0035 and N66001-75-C-0224, and in part by the Office of Naval Research under Grant N00014-75-C-0518.

STOCHASTIC APPROXIMATION

WITH

CORRELATED DATA[1]

by

David C. Farden

Submitted to

The Annals of Statistics

April 1976

[1]This work was supported in part by the Advanced Research Projects Agency of the Department of Defense and monitored by the Naval Undersea Center, San Diego, under Contract Numbers N66001-74-C-0035 and N66001-75-C-0224, and in part by the Office of Naval Research under Grant N00014-75-C-0518.

The author is with the Department of Electrical Engineering, Colorado State University, Fort Collins, Colorado 80523.

American Mathematical Society 1970 subject classifications. Primary 61L20, 62M10; Secondary 62J05, 60G35, 94A05.

Key words and phrases. Stochastic approximation, Robbins-Monro procedure, moment conditions, autocovariance decay rate conditions, almost sure convergence.

SUMMARY

STOCHASTIC APPROXIMATION WITH CORRELATED DATA

New almost sure convergence results are developed for a special form of the multidimensional Robbins-Monro (RM) stochastic approximation procedure. The spe ial form treated can be viewed as a stochastic approximation to the solution $w = w_o \varepsilon R^p$ of the linear equations $Rw = P$, where R is a pxp positive definite symmetric matrix. This special form commonly arises in adaptive signal processing applications. Essentially, previous convergence results for the RM procedure contain a common "conditional expectation condition" which is extremely difficult (if not impossible) to satisfy when the "training data" is a correlated sequence. In contrast, the new convergence results incorporate moment conditions and covariance function decay rate conditions. The ease with which these results can be applied in many cases is illustrated.

1. Introduction. Consider the set of linear equations $Rw=P$, where $R$ is a pxp symmetric positive definite matrix, and $w$ and $P$ are px1 matrices. In case $R$ and $P$ are unknown, and the solution, $w = w_o = R^{-1}P$, is desired, many techniques are available for finding an estimate of $w_o$. In many adaptive signal processing applications, a recursive, computationally efficient procedure for estimating $w_o$ is an important issue. A suitable multidimensional version of the Robbins-Monro (RM) stochastic approximation procedure (Robbins and Monro (1951)) for recursively estimating $w_o$ is given by

(1) $W_{n+1} = W_n + \mu_n(P_n - F_n W_n), \ n \geq 1,$

where $\{P_n\}$ is a sequence of random variables, $P_n \epsilon R^p$, $\{F_n\}$ is a sequence of random pxp matrices, $\{\mu_n\}$ is a sequence of positive constants, and $W_1 \epsilon R^p$ is arbitrary. It is assumed that $P_n$ and $F_n$ are functionally independent of $W_1, W_2, \ldots, W_n$. It is somewhat helpful to consider $w_o$ to be the vector which minimizes $\xi(w) = w'Rw - 2w'P$, where $'$ denotes matrix transpose. Interpreting $F_n$ and $P_n$ as "instantaneous estimates" of $R$ and $P$, respectively, the relationship between (1) and (deterministic) steepest descent procedures is obvious. Consequently, the family of algorithms represented by (1) has an interpretation as a family of "stochastic gradient-following" algorithms.

Algorithm (1) provides a suitable framework for the analytical treatment of many of the algorithms that have been proposed in the engineering literature for adaptive signal processing applications (e.g., see Sakrison (1966) or Farden (1975)). For such applications, any "conditional expectation assumption" is extremely difficult (if not impossible) to establish. Such conditions are commonly required by existing convergence theorems treating the RM procedure. For example, see Schmetterer (1961, 1969), Sakrison (1966), or Farden (1975) for a discussion of existing results. The practical application of existing convergence results for the RM procedure to (1) essentially requires that $\{P_n - F_n w\}$ is an uncorrelated sequence for all fixed parameter $w \varepsilon R^p$. The special form of the RM procedure represented by (1) enables us to obtain convergence results which make use of no such conditional expectation requirements and have a decidedly different flavor than existing results for the RM procedure.

The contents and organization of this paper are as follows. Notation and basic assumptions are presented in Section 2. The framework presented in Assumption (2.1) establishes that the sequences $\{F_n\}$ and $\{P_n\}$ are such that the "time averages" of $E(F_n)$ and $E(P_n)$, respectively, are equal to R and P. The generalization resulting from these definitions is applicable to cases where $E(F_n)$ and $E(P_n)$ are periodic, such as occurs in some adaptive digital communication applications. In Section 3, it is shown that Assumptions (2.1) through (2.7) are sufficient for the a.s. convergence of $W_n$ to $w_o$. The proofs of Lemma (3.1) and Theorem (3.2) below are quite similar in spirit to the proofs of Theorem (6.1) and Theorem (6.3) of Albert and Gardner (1967), respectively. However, the seemingly less restrictive assumptions made in the present work, the simplification in proof resulting for symmetric $F_n$, and the basic differences in the form of algorithms treated here are offered as justification for including the results of Section 3 in this paper. Furthermore, in contrast with the assumptions made by Albert and Gardner (1967), the form of Assumptions (2.3) - (2.7) permits us to exploit the Borel-Cantelli Lemma and the results of Serfling (1970) to prove Corollary (4.5). Corollary (4.5) provides easily verified sufficient conditions for Assumptions (23)-(2.7), and hence, for the a.s. convergence of $W_n$ to $w_o$. Several special cases of (1) are considered in Section 5 to illustrate the application of these results. In case $F_n$ and $P_n$ are strongly consistent estimates of R and P, respectively, the much simpler convergence result of Section 6 is applicable.

2. Notation and basic assumptions. The norm of a pxp matrix A, denoted by $||A||$, is defined here by $||A|| = \sup_{|x|=1} |Ax|$, $x \in R^p$, where $R^p$ denotes p-dimensional Euclidean space and $|x|$, $x \in R^p$ denotes the usual p-dimensional Euclidean norm. For A real and symmetric, $||A|| = \max_i \{|\lambda_i(A)|\}$, where $\{\lambda_i(A)\}_{i=1}^p$ are the p eigenvalues of A. The minimum and maximum eigenvalues of a pxp matrix A are denoted by $\lambda_{min}(A)$ and $\lambda_{max}(A)$, respectively. The element of a pxp matrix A occurring in the ith row and jth column of A is denoted by $(A)_{i,j}$. Similarly, the ith element of $x \in R^p$ is denoted by $(x)_i$. The trace of a pxp matrix A is denoted by $tr(A) = \sum_{i=1}^p (A)_{i,i}$. The symbol 0 is used to denote the additive identity for $R^1, R^p$, or to denote a pxp matrix of zeros. Square brackets [ ] are used to denote integer part. Finally, subscripted variables like $\nu_k$, $n_\ell$, $\ell_i$, etc. are sometimes denoted by $\nu(k)$, $n(\ell)$, $\ell(i)$, etc.

All random variables are assumed to be defined on a probability space $(\Omega, F, P)$. All relationships between random variables are to be interpreted to hold with probability one.

It is worth emphasizing that Corollary (4.5) below establishes sufficient conditions for Assumptions (2.3)-(2.7).

(2.1) ASSUMPTION. The sequence $\{W_n\}_{n=1}^\infty$, $W_n \in R^p$, satisfies the recursion

$$(1) \quad W_{n+1} = W_n + \mu_n(P_n - F_n W_n), \; n \geq 1,$$

where $\{P_n\}$ is a sequence of random variables, $P_n \in R^p$, $\{F_n\}$ is a sequence of real symmetric nonnegative definite (pxp) random matrices, $\{\mu_n\}$ is a sequence of positive real numbers, and $W_1 \in R^p$ is arbitrary.

Define $w_o = R^{-1}P$, where $R = \lim_{n\to\infty} n^{-1}\sum_{\ell=1}^n E(F_\ell)$, $P = \lim_{n\to\infty} n^{-1}\sum_{\ell=1}^n E(P_\ell)$. It is assumed that $E(F_\ell)$, $E(P_\ell)$ exist, that the above limits exist, and that R is positive definite. Further, it is assumed that $n^{-1}\sum_{\ell=a+1}^{n+a} E(F_\ell)$ converges uniformly to R as $n\to\infty$ for all non-negative integers a.

Defining $V_n = W_n - w_o$ and $C_n = P_n - F_n w_o$, we have

(2) $V_{n+1} = (I - \mu_n F_n)V_n + \mu_n C_n$ .

For a sequence $\{A_i\}$ of pxp matrices, define

(3) $$\prod_{i=\ell}^{k} A_i = \begin{cases} A_k A_{k-1} \cdots A_\ell, & \text{if } k \geq \ell \\ I, & \text{if } k < \ell. \end{cases}$$

Defining $Q_{\ell,m} = \prod_{j=\ell}^{m} (I - \mu_j F_j)$, $\Lambda_n = \sum_{k=1}^{n} Q_{k+1,n}\mu_k C_k$,

and iterating (2), one obtains

(4) $V_{n+1} = Q_{1,n}V_1 + \Lambda_n$.

(2.2) ASSUMPTION. The sequence $\{\mu_n\}_1^\infty$ is a nonincreasing sequence of positive constants $\mu_n = O(n^{-1})$, $0 < \lim_{n\to\infty} n\mu_n < \infty$.

(2.3) ASSUMPTION. Assumptions (2.1) and (2.2) hold and $\mu_n||F_n|| \xrightarrow{a.s.} 0$ as $n \to \infty$.

(2.4) ASSUMPTION. Assumption (2.1) holds and $n^{-1}\sum_{k=a+1}^{a+n} F_k \xrightarrow{a.s.} R$ as $n \to \infty$ for all positive integers a.

(2.5) ASSUMPTION. There exists a sequence of random integers $\{\nu_k\}$ with $1 = \nu_1 < \nu_2 < \nu_3 < \ldots$ such that, with $p_k = \nu_{k+1} - \nu_k$ and $J_k = \{\nu_k, \nu_k+1, \ldots, \nu_{k+1}-1\}$ we have (i) $p_k = \xi + [k^\alpha]$, for some $\alpha$, $0 \leq \alpha < 1$, (ii) $p_k^{-1} \lambda_{min}(\sum_{j\in J_k} F_j) \geq \delta > 0$, and (iii) $p_k^{-1}\lambda_{max}(\sum_{j\in J_k} F_j) < \gamma < \infty$. The quantities $\xi$, $\delta$, and $\gamma$ are all random variables that are independent of k.

(2.6) ASSUMPTION. Assumptions (2.1) and (2.2) hold and there exists a random variable $S \in R^p$ such that $S_n \triangleq \sum_{k=1}^{n} \mu_k C_k \xrightarrow{a.s.} S$ as $n\to\infty$.

(2.7) ASSUMPTION. Assumptions (2.1) and (2.6) hold and $|F_n(S - S_{n-1})| \xrightarrow{a.s.} 0$ as $n\to\infty$.

3. <u>Almost sure convergence of $W_n$ to $w_o$.</u>

(3.1) LEMMA. *If assumptions (2.1)-(2.5) are satisfied, then* $||Q_{1,n}|| \xrightarrow{a.s.} 0$.

PROOF. For any positive integer n, let $K = K(n)$ be such that $n \in J_K$. Then $Q_{1,n} = Q_{\nu(K),n} Q_{1,\nu(K)-1}$, and hence $||Q_{1,n}|| \le ||Q_{\nu(K),n}|| \cdot ||Q_{1,\nu(K)-1}||$. Defining $\Gamma_k = Q_{\nu(k),\nu(k+1)-1}$, $\Gamma_k$ may be expressed as

$$\Gamma_k = I - \sum_{j \in J_k} \mu_j F_j + \sum_{q=2}^{p_k} (-1)^q \sum_{L_{k,q}} \prod_{i=1}^{q} \mu_{\ell(i)} F_{\ell(q-i+1)},$$

where $L_{k,q} = \{\ell_i \in J_k : \ell_1 > \ell_2 > \ldots > \ell_q, i=1,2,\ldots,q\}$,

so that (for $\mu_{\nu(k+1)-1} p_k \delta < 1$)

$$||\Gamma_k|| \le 1 - \mu_{\nu(k+1)-1} p_k \delta + \sum_{q=2}^{p_k} (\mu_{\nu(k)} p_k \gamma)^q = 1 - \mu_{\nu(k+1)-1} p_k \delta + \frac{\alpha_k^2 - \alpha_k^{p_k+1}}{1 - \alpha_k},$$

where $\alpha_k = \mu_{\nu(k)} p_k \gamma$ and the last equality holds provided that $\alpha_k \neq 1$. Assumptions (2.2) and (2.5) imply that there exists a random positive integer $k_o$ such that $\alpha_k \le \frac{1}{2}$ and $||\Gamma_k|| \le 1 - \beta_k$ for all $k \ge k_o$, where $\beta_k = \frac{1}{2} \mu_{\nu(k+1)-1} p_k \delta$. Similarly, for all $K \ge k_o$ and $n \in J_K$, $||Q_{\nu(K),n}|| \le 3/2$. It follows that there exists a random variable M such that for all $K \ge k_o$,

$$||Q_{1,n}|| \le \frac{3}{2} M \prod_{k=k_o}^{K} (1 - \beta_k) \le \frac{3}{2} M \exp\left(-\frac{1}{2}\delta \sum_{k=k_o}^{K} \mu_{\nu(k+1)-1} p_k\right),$$

since $1 - x \le e^{-x}$ for all real x. It is easily shown that Assumptions (2.2) and (2.5) imply that the above summation diverges to $\infty$ as $K \to \infty$. Consequently, $||Q_{1,n}|| \xrightarrow{a.s.} 0$ as $n \to \infty$. □

(3.2) THEOREM. *If Assumptions (2.1)-(2.7) are satisfied, then* $|V_n| \xrightarrow{a.s.} 0$ *as* $n \to \infty$.

PROOF. From (2.1.4) and Lemma (3.1), it remains only to show that $|\Lambda_n| \xrightarrow{a.s.} 0$ as $n \to \infty$. From Assumptions (2.1) and (2.6), with $S_0 = 0$ and $Q_{n+1,n} = I$, we have

$$(1) \quad \Lambda_n = \sum_{k=1}^{n} (Q_{k,n} - Q_{k+1,n}) S_{k-1} + S_n$$

$$= - \sum_{k=1}^{n} Q_{k+1,n} \mu_k F_k S_{k-1} + S_n \quad .$$

Defining

$$B_{\ell,n} = \sum_{k=\ell}^{n} Q_{k+1,n} \mu_k F_k (S-S_{k-1}),$$

and

$$D_n = \sum_{k=1}^{n} Q_{k+1,n} \mu_k F_k S,$$

(1) may be expressed as $\Lambda_n = B_{1,n} - D_n + S_n$.

From Lemma 1 of Albert and Gardner (1967, p. 189), $D_n = (I-Q_{1,n})S$, so that $\Lambda_n = B_{1,n} + (S_n-S) + Q_{1,n} S$. Since $S_n \xrightarrow{a.s.} S$ and $||Q_{1,n}|| \xrightarrow{a.s.} 0$ as $n \to \infty$ (Assumption (2.6) and Lemma (3.1)), it remains to show that $|B_{1,n}| \xrightarrow{a.s.} 0$ as $n \to \infty$.

Using the same notation as in the proof of Lemma (3.1), $|B_{1,n}|$ may be bounded as

$$|B_{1,n}| \leq ||Q_{\nu(K),n}|| \cdot | \sum_{k=k_o}^{K-1} \prod_{i=k+1}^{K-1} \Gamma_i \sum_{j\varepsilon J_k} Q_{j+1,\nu(k+1)-1} \mu_j F_j (S-S_{j-1}) |$$

$$+ ||Q_{\nu(K),n}|| \prod_{\ell=k_o}^{K-1} (1-\beta_\ell) | \sum_{k=1}^{k_o-1} \prod_{i=k+1}^{k_o-1} \Gamma_i \sum_{j\varepsilon J_k} Q_{j+1,\nu(k+1)-1} \mu_j F_j (S-S_{j-1}) |$$

$$+ | \sum_{k=\nu(K)}^{n} Q_{k+1,n} \mu_k F_k (S-S_{k-1}) | .$$

Note that for all $j,m \varepsilon J_k$ and $k \geq k_o$ we have $||Q_{j,m}|| \leq 1 + \sum_{q=2}^{p_k} (\mu_{\nu(k)} p_k \gamma)^q \leq \frac{3}{2}$.

Consequently, defining $d_k = \max_{j\varepsilon J_k} |F_j (S-S_{j-1})|$, there exists a random variable $M_1$ such that

$$|B_{1,n}| \le \frac{9}{4} \sum_{k=k_o}^{K-1} \prod_{i=k+1}^{K-1} (1-\beta_i)\, p_k\, \mu_{\nu(k)} d_k$$

$$+ \frac{3}{2} M_1 \prod_{i=k_o}^{K-1} (1-\beta_i) + \frac{3}{2} p_K \mu_{\nu(K)} d_K \ .$$

Defining $a_{n,i} = \prod_{\ell=i+1}^{n} (1-\beta_\ell)\beta_i$, it remains only to show that

$\sum_{k=k_o}^{K-1} a_{K-1,k}\beta_k^{-1}\, p_k\, \mu_{\nu(k)} d_k \xrightarrow{a.s.} 0$ as $K \to \infty$. Clearly, for all fixed $i \ge k_o$, $a_{n,i} \xrightarrow{a.s.} 0$ as $n \to \infty$. From Lemma 1 of Albert and Gardner (1967, p. 189), $\sum_{i=k_o}^{n} |a_{n,i}| = 1 - \prod_{i=k_o}^{n} (1-\beta_i)$, which converges a.s. to 1 as $n \to \infty$. Consequently, by the Toeplitz Lemma (e.g., see Knopp (1947, p. 75)), $\lim_{K\to\infty} \sum_{k=k_o}^{K-1} a_{K-1,k}\, \mu_{\nu(k)} d_k p_k \beta_k^{-1} = \lim_{k\to\infty} \mu_{\nu(k)} d_k p_k \beta_k^{-1}$. By Assumptions (2.2) and (2.5), $\mu_{\nu(k)} p_k \beta_k^{-1} = \frac{2}{\delta} \frac{\mu_{\nu(k)}}{\mu_{\nu(k+1)-1}}$ is bounded; hence, $|B_{1,n}| \xrightarrow{a.s.} 0$ as $n \to \infty$. □

(3.3) COROLLARY. *If $\{\mu_k\}$ satisfies Assumption (2.2) and $||F_k||$ is a.s. uniformly bounded (in k), then Assumptions (2.3) and (2.7) may be deleted and Theorem (3.2) remains true.*

PROOF. It suffices to consider $\mu_k = k^{-1}$. Since $|F_n(S-S_{n-1})| \le ||F_n|| \cdot |S-S_{n-1}|$, Assumption (2.6) implies Assumption (2.7). The Borel-Cantelli Lemma and the Chebychev inequality can easily be applied to show that Assumption (2.3) is satisfied. □

4. Sufficient conditions for Assumptions (2.3) - (2.7). Several auxiliary lemmas will be needed before the main result of this section, Corollary (4.5) may be stated and proved. The following lemma is a reasonably straightforward extension of Theorem A presented by

Serfling (1970). Consequently, the proof will be omitted.

(4.1) LEMMA. *Let* $\{x_i\}$ *be a sequence of random variables,* $x_i \varepsilon R^p$, *having finite "variances"* $\sigma_i^2 = E[(x_i - E(x_i))'(x_{a+n}{}_i - E(x_i))]$. *For integer* $n \geq 1$ *define* $X_{a,n} = (x_{a+1}, \ldots, x_{a+n})$, $S_{a,n} = \sum_{i=a+1} x_i$, *and* $M_{a,n} = \max \{|S_{a,1}|, \ldots, |S_{a,n}|\}$. *For* $n \leq -1$ *define* $X_{a,n} = (x_{a+n+1}, \ldots, x_a)$, $S_{a,n} = \sum_{i=a+n+1}^{a} x_i$, *and* $M_{a,n} = \max \{|S_{a,n}|, \ldots, |S_{a,-1}|\}$. *For* $|n| \geq 1$ *let* $F_{a,n}$ *denote the distribution function for* $X_{a,n}$, *and let* $g(F_{a,n})$ *be a functional depending on* $F_{a,n}$. *Let* $a_o$ *be an arbitrary but fixed integer and let* $\nu \geq 2$. *Suppose* $g(F_{a,k}) + g(F_{a+k,\ell}) \leq g(F_{a,k+\ell})$ *for all* $a \geq a_o$ *and* $1 \leq k \leq k + \ell$ *or* $a_o - a \leq k + \ell \leq k \leq -1$ *such that* $E(|S_{a,n}|^\nu) \leq g^{\frac{1}{2}\nu}(F_{a,n})$ *for all* $a \geq a_o$ *and* $n \geq 1$ *or* $a_o - a \leq n \leq -1$. *Then* $E(M_{a,n}^\nu) \leq (\log_2 2|n|)^\nu g^{\frac{1}{2}\nu}(F_{a,n})$ *for all* $a \geq a_o$ *and* $n \geq 1$ *or* $a_o - a \leq n \leq -1$.

(4.2) LEMMA. *Let* $\alpha_{k,\ell} = \alpha_{\ell,k}$ *and* $\rho(k,\ell) = \rho(\ell,k)$ *be real-valued functions defined for all non-negative integers* $k, \ell$. *For* $1 \leq n < m$ *define*

$$(1) \quad \gamma_{n,m} = \sum_{k=n}^{m} \sum_{\ell=n}^{m} \alpha_{k,\ell}\rho(k,\ell)$$

$$= 2 \sum_{u=1}^{m-n} \sum_{k=n}^{m-u} \alpha_{k,k+u}\, \rho(k,k+u) + \sum_{k=n}^{m} \alpha_{k,k}\, \rho(k,k).$$

*Suppose that* $|\rho(k,k+u)| = O(u^{-\nu})$ *uniformly for all positive integers* $k$. *If* $\alpha_{k,\ell} = 1$ *and* $0<\nu<1$, *then for large* $m-n$,

$$(2) \quad |\gamma_{n,m}| = O((m-n)^{2-\nu}).$$

*Finally, if* $\alpha_{k,\ell} = \mu_k\mu_\ell$, $\mu_k = O(k^{-1})$, *and* $0<\nu<1$, *then*

$$(3) \quad |\gamma_{n,m}| = O(n^{-\nu/(\nu+2)}).$$

PROOF. Suppose that $\alpha_{k,\ell} = 1$ and $|\rho(k,k+u)| = O(u^{-\nu})$, $0<\nu<1$, uniformly for all positive integers k. It suffices to consider $|\rho(k,k+u)| = (|u|+1)^{-\nu}$ in which case (from (1))

$$|\gamma_{n,m}| \leq 2(m-n+1) \sum_{u=0}^{m-n} (u+1)^{-\nu} \leq 2(m-n+1)(1+\int_1^{m-n+1} x^{-\nu} dx).$$

The result (2) follows easily by evaluating the above integral.

Suppose now that $\alpha_{k,\ell} = \mu_k \mu_\ell$, $\mu_k = O(k^{-1})$, and $|\rho(k,k+u)| = O(u^{-\nu})$, $0<\nu<1$, uniformly for all positive integers k. For this case, it suffices to consider $\mu_k = k^{-1}$ and $|\rho(k,k+u)| = (|u|+1)^{-\nu}$. Then from (1)

$$|\gamma_{n,m}| \leq 2 \sum_{u=1}^{m-n} (u+1)^{-\nu} \int_{n-1}^{m-u} x^{-1}(x+u)^{-1} dx + (n-1)^{-1} - m^{-1}$$

$$= 2 \sum_{u=1}^{m-n} u^{-1}(u+1)^{-\nu} \beta_{n,m,u} + (n-1)^{-1} - m^{-1},$$

where $\beta_{n,m,u} = \ell n\ ((m-u)(n+u-1)) - \ell n(m(n-1))$. If $1 \leq u \leq \ell \leq n < m$, then $\beta_{n,m,u} \leq \ell n(n+\ell-1) - \ell n(n-1)$. If $1 \leq \ell < u < n < m-n$, then $\beta_{n,m,u} \leq \ell n\ 2$. If $n \leq u \leq m-n$, then $\beta_{n,m,u} \leq \ell n(u)$. Consequently, for all $1 \leq \ell \leq n-2 < m-n-2$,

$$|\gamma_{n,m}| \leq 2\ell\ \ell n \left(\frac{n+\ell-1}{n-1}\right) + 2\ \ell n\ 2 \sum_{u=\ell+1}^{n-1} u^{-1-\nu}$$

$$+ 2 \sum_{u=n}^{m-n} u^{-1-\nu} \ell n\ u + (n-1)^{-1} - m^{-1}.$$

Letting $\ell = n^{\beta}$, $1>\beta>0$, and using the fact that $\ell n\ (1+x) \leq x$ for all $x \geq -1$, it follows that there exist constants $c_1$, $c_2$, and $c_3$ such that

$$|\gamma_{n,m}| \leq c_1 n^{2\beta-1} + c_2 n^{-\beta\nu} + c_3 n^{-\nu/2}.$$

Finally, if $\beta = (\nu+2)^{-1}$, then $|\gamma_{n,m}| = O(n^{-\nu/\nu+2})$. □

(4.3) LEMMA. *Define*

$$\gamma_F(i,j,\ell,m) = \max_{|w|=1} w'E((F_i-E(F_i))ww'(F_j-E(F_j))ww'(F_\ell-E(F_\ell))ww'(F_m-E(F_m)))w,$$

*where* $w\epsilon R^p$. *Define* $P_k(a) = a + [k^\alpha]$, *where* $a$ *is a positive integer,* $0 \leq \alpha < 1$. *Define the sequence* $\{\nu_k(a)\}$ *by* $\nu_1(a) = 1$, $\nu_{k+1}(a) = \nu_k(a) + P_k(a)$, $k = 1,2,\ldots$, *and define* $J_k(a) = \{\nu_k(a), \nu_k(a)+1, \ldots, \nu_{k+1}(a)-1\}$. *If Assumption (2.4) holds and*

$$(1) \quad \sum_{k=1}^{\infty} P_k^{-4}(a) \sum_{i,j,\ell,m\epsilon J_k(a)} \gamma_F(i,j,\ell,m) < \infty ,$$

*for some* $\alpha$, $0 \leq \alpha < 1$, *and for some positive integer* $a$, *then Assumption (2.5) is satisfied.*

PROOF. Define $G_k(a) = P_k^{-1}(a) \sum_{j\epsilon J_k(a)} F_j$ and $R_k(a) = P_k^{-1}(a) \sum_{j\epsilon J_k(a)} E(F_j)$.

Let $\varepsilon$ be given such that $0 < \varepsilon < \lambda_{\min}(R)$. Since $\lambda_{\min}(G_k(a)) \geq \lambda_{\min}(G_k(a)-R) + \lambda_{\min}(R)$ and $\lambda_{\max}(G_k(a)) \leq \lambda_{\max}(G_k(a)-R) + \lambda_{\max}(R)$, it is sufficient to show that there exists a random sequence $\{\nu_k(\xi)\}$, $\xi$ an integer-valued random variable with $\xi$ a.s. finite such that

$$0<\lambda_{\min}(R) - \varepsilon < \lambda_{\min}(G_k(\xi)-R) + \lambda_{\min}(R) \leq \lambda_{\max}(G_k(\xi)-R) + \lambda_{\max}(R)$$
$$<\lambda_{\max}(R) + \varepsilon$$

for all k, or, equivalently, that

$\max_{|w| = 1} |w'(G_k(\xi)-R)w| < \varepsilon$ for $k = 1,2,\ldots$. Hence, it is also sufficent to require the stronger condition that

$$\max_{|w|= 1}|w'(G_k(\xi)-R_k(\xi))w| + \max_{|w|= 1}|w'(R_k(\xi)-R)w| < \varepsilon$$

for all k . By Assumption (2.1), there exists a positive integer $q_1$ such that for all sequences $\{\nu_k(a)\}$ with $a \geq q_1$ we have

$\max\limits_{|w|=1} |w'(R_k(a)-R)w| < \varepsilon/2$. It follows from Assumption (2.4) that for all $n=1,2,\ldots$, there exists an a.s. finite random variable $\xi_n^*$ such that

$$\max_{1\le k\le n} \max_{|w|=1} |w'(G_k(\xi_n^*)-R_k(\xi_n^*))w| < \varepsilon/2.$$

Let $\xi_n$ be the smallest such $\xi_n^*$ so that $\xi_n \ge q_1$.

It follows that

$$P(\xi_n \ge a \ge q_1) \le P\left(\max_{|w|=1} |w'(G_k(a)-R_k(a))w| \ge \varepsilon/2 \text{ for some } k,\ 1\le k\le n\right)$$

$$\le \sum_{k=1}^{n} P\left(\max_{|w|=1} |w'(G_k(a)-R_k(a))w| \ge \varepsilon/2\right)$$

$$\le \sum_{k=1}^{n} \left(\frac{\varepsilon}{2}P_k(a)\right)^{-4} E\left(\max_{|w|=1} |w'(G_k(a)-R_k(a))w|^4\right)$$

$$= \left(\frac{2}{\varepsilon}\right)^4 \sum_{k=1}^{n} (P_k(a))^{-4} \sum_{i,j,\ell,m\in J_k(a)} \gamma_F(i,j,\ell,m).$$

By hypothesis the above series converges as $n \to \infty$. Applying the definition of $P_k(a)$, it follows that for all $\varepsilon_1 > 0$, there exists a positive integer $N(\varepsilon_1)$ such that for all $a \ge N(\varepsilon_1)$ we have $\lim\limits_{n\to\infty} P(\xi_n \ge a \ge q_1) < \varepsilon_1$. Consequently, $\xi = \sup\limits_n \xi_n$ is a.s. finite and Assumption (2.5) is satisfied. $\square$

(4.4) LEMMA. *If there exists a real-valued function $f(k,k+u) = O(|u|^{-\beta})$, $\beta > \frac{1}{4}$, uniformly for all positive integers $k$ such that*

$$|\gamma_F(i,j,\ell,m)| \le f^2(i,j)f^2(\ell,m) + f(i,j)f(i,\ell)f(j,m)f(\ell,m),$$

*then Assumption (2.5) is satisfied.*

PROOF.

It suffices to consider $f(i,j) = (|i-j|+1)^{-\beta}$.

Define the ijth element of a matrix A to be f(i,j). Then

$$\sum_{i=a}^{b}\sum_{j=a}^{b}\sum_{\ell=a}^{b}\sum_{m=a}^{b} f(i,j)f(i,\ell)f(j,m)f(\ell,m) = tr(A^4)\le(trA^2)^2.$$

The iith element of $A^2$ is given by

$$(A^2)_{i,i} = \sum_{m=a}^{i-1} (i-m+1)^{-2\beta} + \sum_{m=i+1}^{b} (m-i+1)^{-2\beta} + 1$$

$$\le c_1(i-a+1)^{1-2\beta} + c_2(b-i+1)^{1-2\beta},$$

for $0<\beta<\frac{1}{2}$ and some positive constants $c_1$ and $c_2$. Consequently, for some constant $c_3$, $tr(A^2) \le c_3(b-a+1)^{2-2\beta}$, and hence $tr(A^4) = O((b-a)^{4-4\beta})$. From Lemma (4.2), $(\sum_{i=a}^{b}\sum_{j=a}^{b} f^2(i,j))^2 = O((b-a)^{4-4\beta})$. Clearly, (4.3.1) be satisfied if $\sum_{k=1}^{\infty} k^{-4\alpha}\, k^{-4\alpha-4\alpha\beta}<\infty$. It follows that Assumption (2.5) is satisfied for $1>\alpha>1/4\beta$ if $\beta>\frac{1}{4}$. □

(4.5) COROLLARY. *Suppose that Assumptions (2.1), (2.2), and (2.5) are satisfied. Define* $\rho_F(k,\ell) = E(F_k F_\ell) - E(F_k)E(F_\ell)$, $\rho_C(k,\ell) = E(C_k' C_\ell) - E(C_k')E(C_k)$, *and* $\rho_{FC}(k,\ell,n) = E((C_k'-E(C_k'))F_n^2(C_\ell-E(C_\ell)))$, *where all expectations are assumed to exist. Suppose that for some* $\nu>0$, $u^\nu \max\{||\rho_F(k,k+u)||, |\rho_C(k,k+u)|, |\rho_{FC}(k,k+u,n)|\}$ *is uniformly bounded for all non-negative integers* $k$, $u$, *and* $n$. *Further suppose that* $g_n = \sum_{k=n}^{\infty} \mu_k E(C_k)$ *exists and that there exists a constant* $\beta > 1$ *such that* $\sum_{n=1}^{\infty} |g_n|^\beta E(||F_n||^\beta)<\infty$. *Then* $|V_n| \xrightarrow{a.s.} 0$ *as* $n\to\infty$.

PROOF. First consider Assumption (2.4). Define $S_{a,n} = \sum_{k=a+1}^{a+n} (F_k - E(F_k))w$, where $w \in R^p$. Clearly, Assumption (2.4) is satisfied iff $n^{-1}|S_{a,n}| \xrightarrow{a.s.} 0$ for all positive integers a as $n \to \infty$. Define $M_{a,n} = \max\{|S_{a,1}|, \ldots, |S_{a,n}|\}$. Let $\{n_k\}$ be an increasing sequence of positive integers such that $n_k \to \infty$ as $k \to \infty$. For all $n_k \leq n \leq n_{k+1}-1$,

$$(1) \quad n^{-1}|S_{a,n}| \leq n_k^{-1}|S_{a,n(k)-1}| + n_k^{-1} M_{a+n(k)-1,n(k+1)-n(k)}.$$

To apply Lemma (4.1), define $g(F_{a,n})$ as

$$(2) \quad g(F_{a,n}) = |w|^2 \sum_{k=a+1}^{a+n} \sum_{\ell=a+1}^{a+n} ||\rho_F(k,\ell)|| = O(n^{2-\nu}),\ 0<\nu<1,$$

uniformly for all positive integers a, from Lemma (4.2). It is easily seen that $g(F_{a,n})$ satisfies the needed conditions of Lemma (4.1). Letting $n_k = k^\alpha$, $n_k^{-2} E(|S_{a,n(k)}|^2)$ is summable for all $\alpha > \nu^{-1}$. The Chebychev inequality and Borel-Cantelli Lemma thus imply that $n_k^{-1}|S_{a,n(k)-1}| \xrightarrow{a.s.} 0$ as $k \to \infty$. Letting $\xi_k = M_{a+n(k)-1,\ n(k+1)-n(k)}$,

$$(3) \quad n_k^{-2} E(\xi_k^2) = O(n_k^{-\nu}(\log_2 2(n_{k+1}-n_k))^2),$$

from (2) and Lemma (4.1). Substituting $n_k = k^\alpha$ into (3), the Borel-Cantelli Lemma and the Chebychev inequality imply that $n_k^{-1}\xi_k \xrightarrow{a.s.} 0$ as $k \to \infty$ for $\alpha > \nu^{-1}$. Consequently, from (1), Assumption (2.4) is satisfied.

Now consider Assumption (2.6). Define $S_{a,m} = \sum_{k=a+m+1}^{a} \mu_k(C_k - E(C_k))$, $M_{a,m} = \max\{|S_{a,m}|, \ldots, |S_{a,-1}|\}$, and $g(F_{a,m}) = \sum_{k=a+m+1}^{a} \sum_{\ell=a+m+1}^{a} \mu_k \mu_\ell |\rho_C(k,\ell)|$. It is easily seen that $g(F_{a,m})$ satisfies the needed conditions of Lemma (4.1). From

Lemma (4.2), $E(|S_{a,n-a-1}|^2) = O(n^{-\nu/(\nu+2)})$ for all $a > n$. With $n_k=k^{\alpha}$, the Chebychev inequality and the Borel-Cantelli Lemma thus imply that

$$S-S_{n(k)-1}-g_{n(k)} = \sum_{i=n_k}^{\infty} \mu_i(C_i - E(C_i)) \xrightarrow{a.s.} 0 \text{ as } k \to \infty \text{ for } \alpha>\nu^{-1}(\nu+2).$$

Consequently, $S-S_{n(k)-1} \xrightarrow{a.s.} 0$ as $k \to \infty$. For all $n_k \leq n \leq n_{k+1}-1$,

$$(4) \quad |S-S_{n-1}| \leq |S-S_{n(k+1)-1}| + \xi_k + b_k,$$

where $\xi_k = M_{n(k+1)-1,n(k)-n(k+1)}$ and $b_k$ is a sequence of positive constants converging to zero as $k \to \infty$. From Lemmas (4.1) and (4.2),

$$(5) \quad E(\xi_k^2) = O((\log_2 2(n_{k+1}-n_k))^2 n_k^{-\nu/(\nu+2)}).$$

Substituting $n_k = k^{\alpha}$ into (5), the Chebychev inequality and the Borel-Cantelli Lemma imply that $\xi_k \xrightarrow{a.s.} 0$ as $k \to \infty$ for all $\alpha > \nu^{-1}(\nu+2)$. Consequently, Assumption (2.6) is satisfied.

Finally, consider Assumption (2.7). Define $Z_{k,n} = F_n(C_k-E(C_k))$,

$$S_{a,m} = \sum_{k=a+m+1}^{a} \mu_k Z_{k,a+m+1}, \text{ and } g(F_{a,m}) = \sum_{k=a+m+1}^{a} \sum_{\ell=a+m+1}^{a} \mu_k\mu_\ell|\rho_{FC}(k,\ell,a+m+1)|.$$

Proceeding as above, $E(|S_{a,n-a-1}|^2) = O(n^{-\nu/(\nu+2)})$ for all $a > n$. With $n_k = k^{\alpha}$, $\alpha > \nu^{-1}(\nu+2)$, we have

$$F_{n(k)}(S-S_{n(k)-1})-F_{n(k)}g_{n(k)} = \sum_{i=n_k}^{\infty} \mu_i Z_{i,n(k)} \xrightarrow{a.s.} 0,$$

as $k\to\infty$. Since $\sum_{n=1}^{\infty} |g_n|^{\beta}E(||F_n||^{\beta})<\infty$, the Markov inequality and the Borel-Cantelli Lemma imply that $F_n g_n \xrightarrow{a.s.} 0$ as $n\to\infty$. Consequently, for all $n_k \leq n \leq n_{k+1}-1$,

$$|F_n(S-S_{n-1})| \leq |F_{n(k+1)}(S-S_{n(k+1)-1})| + \xi_k + \psi_k,$$

where $\xi_k = M_{n(k+1)-1,n(k)-n(k+1)}$ and $\Psi_k \xrightarrow{a.s.} 0$ as $k\to\infty$. Since $\xi_k$ satisifes (5), $\xi_k \xrightarrow{a.s.} 0$ as $k\to\infty$. Consequently, Assumption (2.7) is satisfied.

Finally, $\beta>1$ and $\mu_n = O(n^{-1})$ easily provide that $\mu_n||F_n|| \xrightarrow{a.s} 0$ by the Markov inequality and the Borel-Cantelli Lemma, and hence, Assumption (2.3) is satisfied.

Consequently, from Theorem (3.2), $|V_n| \xrightarrow{a.s.} 0$ as $n\to\infty$. □

(4.6) REMARK. An important ancillary result contained in Corollary (4.5) is that sufficient conditions for the strong consistency of the usual sample covariance function are provided. For example, let $\{x_k\}_{-\infty}^{\infty}$ be a zero mean wide-sense stationary real-valued normal random process and define $\rho_x(v) = E(x_k x_{k+v})$. Consider $F_k = F_k(u) = x_k x_{k+u}$. It is easily shown that for this case, $\rho_F(k,k+v) = |\rho_x^2(v)+\rho_x(v+u)\rho_x(v-u)|$. The proof of Corollary (4.5) shows that if $\rho_x(u)=O(u^{-\nu})$ for $\nu>0$, then $n^{-1}\sum_{k=1}^{n} x_k x_{k+v} \xrightarrow{a.s.} \rho_x(v)$ as $n \to \infty$. A similar result, presented as Theorem 8B of Parzen (1961), states that $n^{-1}\sum_{k=1}^{n-|v|} x_k x_{k+|v|} \xrightarrow{a.s.} \rho_x(v)$ as $n \to \infty$ provided that there exist positive constants c, q such that $n^{-1}\sum_{u=0}^{n-1} \rho_x^2(u) \le cn^{-q}$ for all positive integers n. The conditions of Parzen (1961) are clearly satisfied by $\rho_x(u) = O(u^{-\nu})$ and $\nu>0$.

(4.7) REMARKS. Recall that sufficient conditions for Assumption (2.5) have been presented as Lemmas (4.3) and (4.4). Lemma (4.4) is useful for several specific choices of $\{F_n\}$, as shown below in Section 5. In case $\{||F_n||\}$ is bounded, or if $\{F_n\}$ and/or $\{P_n\}$ are deterministic, then the conditions of Corollary (4.5) are simplified, as shown below in Corollaries (4.8)-(4.11).

(4.8) COROLLARY. *Let $\rho_F$, $\rho_C$, and $g_n$ be as in Corollary (4.5). Suppose that Assumptions (2.1), (2.2), and (2.5) are satisfied, and that $\{||F_n||\}$ is bounded. If there exists a $\nu>0$ such that $u^{\nu}$ max $\{||\rho_F(k,k+u)||, |\rho_C(k,k+u)|\}$ is uniformly bounded for all non-negative* *integers k and u, and there exists a* $\beta>0$ *such that* $\sum_{n=1}^{\infty} |g_n|^{\beta}<\infty$, *then* $|V_n| \xrightarrow{a.s.} 0$ *as* $n\to\infty$.

PROOF. Simply apply Corollary (3.3) to Corollary (4.5). □

(4.9) COROLLARY. *Suppose that* $F_n$ *is deterministic, and that Assumptions (2.1) and (2.2) are satisfied. Define* $\rho_P(k,\ell)=E(P_k'P_\ell) - E(P_k')E(P_\ell)$, *and* $g_n = \sum_{k=n}^{\infty} \mu_k(E(P_k)-F_k w_o)$. *If there exists a* $\nu>0$ *such that* $u^{\nu}|\rho_P(k,k+u)|$ *is uniformly bounded for all non-negative integers k and u, and if* $g_n \to 0$ *as* $n\to\infty$ *then* $|V_n| \xrightarrow{a.s.} 0$ *as* $n\to\infty$.

PROOF. An obvious consequence of Corollary (4.5). □

(4.10) COROLLARY. *Suppose that* $\{P_n\}$ *is deterministic, and that Assumptions (2.1), (2.2), and (2.5) are satisfied. Define* $\rho_{FF}(k,\ell,n) = E((F_k-E(F_k))F_n^2(F_\ell-E(F_\ell)))$, *and* $g_n = \sum_{k=n}^{\infty} \mu_k(P_k-E(F_k)\, w_o)$. *If there exists a* $\nu>0$ *such that* $u^{\nu} \max\{||\rho_F(k,k+u)||, ||\rho_{FF}(k,k+u,n)||\}$ *is uniformly bounded for all non-negative integers k, u, and n, and there exists a* $\beta>1$ *such that* $\sum_{n=1}^{\infty} |g_n|^{\beta}E(||F_n||^{\beta})<\infty$, *then* $|V_n| \xrightarrow{a.s.} 0$ *as* $n\to\infty$.

PROOF. Follows directly from Corollary (4.5). □

(4.11) COROLLARY. *Suppose that both* $\{F_n\}$ *and* $\{P_n\}$ *are deterministic, and that Assumptions (2.1) and (2.2) are satisfied. If* $\sum_{k=1}^{\infty} \mu_k (P_k-F_k w_o)$ *exists, then* $|V_n| \to 0$ *as* $n\to\infty$.

PROOF. Trivial case of Corollary (4.5). □

5. <u>Application of Corollary 4.5.</u>

(5.1) Special families of $F_n$ and $P_n$. Let $\{X_j\}_{-\infty}^{\infty}$ be a sequence of $R^p$-valued zero-mean random variables and let $\{s_j\}_{-\infty}^{\infty}$ be a sequence of real-valued zero-mean random variables. Define $R_{xx}(k,\ell) = E(X_k X_\ell')$,

$P_s(k,\ell) = E(s_k X_\ell)$ and $\rho_s(k,\ell) = E(s_k s_\ell)$. Suppose that $R_{xx}(k,k+u)$, $P_s(k,k+u)$, and $\rho_s(k,k+u)$ are periodic in k with period N. Define $R = N^{-1} \sum_{k=1}^{N} R_{xx}(k,k)$ and $P = N^{-1} \sum_{k=1}^{N} P_s(k,k)$. It will become apparent in what follows that R and P satisfy Assumption (2.1).

Suppose that it is desired to choose $w \varepsilon R^p$ to minimize $\xi(w) = \sum_{k=1}^{N} E((s_k - w' X_k)^2)$. Such problems arise frequently in adaptive transversal filter channel equalization in digital communications. When N = 1, the problem reduces to the use of jointly wide-sense stationary sequences $\{s_j\}$ and $\{X_j\}$. Assuming that R is positive definite, the desired solution, $w_o$, is given by $w_o = R^{-1}P$. Assume now that R and/or P are unknown, and that it is desired to use algorithm (2.1.1), with $F_n$ and $P_n$ functions of the observed time series, $\{X_j\}$ and $\{s_j\}$, to recursively estimate $w_o$. Obvious candidates for $F_n$ and $P_n$ are

(1) $$F_n = K_n^{-1} \sum_{j=n-K_n+1}^{n} X_j X_j',$$

and

(2) $$P_n = K_n^{-1} \sum_{j=n-K_n+1}^{n} s_j X_j,$$

where $K_n$ is a positive integer; e.g. $K_n = 1$, K, or n. In fact, algorithms represented here by (2.1.1) with $F_n$ and $P_n$ given by (1) and (2) have frequently appeared in the engineering literature for consideration in a wide range of applications.

Note that if $K_n = K$ (a constant), then $E(F_n)$, $E(P_n)$, and $E(C_n)$ are all periodic (in n) with period N. Furthermore, for any $n > 0$, $N^{-1} \sum_{k=n}^{n+N} E(C_k) = 0$. If, in addition $\{\mu_k\}$ satisfies either $\mu_k = a([\frac{k}{N}] + b)^{-1}$ or $\mu_k = a(k+b)^{-1}$, where $a > 0$ and $b \geq 0$, then it can be shown that $|g_n| = |\sum_{k=n}^{\infty} \mu_k E(C_k)| = O(n^{-1})$. If, for example, $E(||F_n||^2)$ is bounded (in n), then $\sum_{n=1}^{\infty} |g_n|^2 E(||F_n||^2) < \infty$, thus satisfying the conditions on $\{g_n\}$stated in Corollaries (4.5) and (4.8)-(4.10). Of course, many other choices of $\{\mu_k\}$ are permissible. Finally, if $K_n = N$, then $E(C_k) \equiv 0$, and the conditions on $\{g_n\}$ are *a fortiori* satisfied. The remaining conditions on $\{F_n\}$ and $\{P_n\}$ stated in the preceding corollaries are quite mild "asymptotic covariance decay rate" conditions. The strongest of these decay rate conditions is that imposed on $\gamma_F$ via Lemma (4.3) in order to satisfy Assumption (2.5).

Regarding the covariance decay rate conditons in the corollaries of Section 4, it may be helpful to note that

$$(3) \quad \rho_C(k,\ell) = \rho_P(k,\ell) + w_o'\rho_F(k,\ell)w_o - \rho_{PF}(k,\ell)\, w_o - \rho_{PF}(\ell,k)w_o \;,$$

where $\rho_P$ and $\rho_F$ are defined in Corollaries (4.9) and (4.5), respectively, and $\rho_{PF}(k,\ell) = E(P_k'F_\ell) - E(P_k')E(F_\ell)$. Furthermore, $\rho_{FC}$ can be expressed as

$$(4) \quad \rho_{FC}(k,\ell,n) = \rho_{PFP}(k,\ell,n) + w_o' \,\rho_{FF}(k,\ell,n)w_o$$

$$-\rho_{PFF}(k,\ell,n)w_o - \rho_{PFF}(\ell,k,n)w_o,$$

where $\rho_{FF}$ is defined in Corollary (4.10) and $\rho_{PFP}(k,\ell,n) = E((P_k'-E(P_k'))\cdot F_n^2(P_\ell-E(P_\ell)))$, and $\rho_{PFF}(k,\ell,n) = E((P_k'-E(P_k'))F_n^2(F_\ell-E(F_\ell)))$. The results (3) and (4) simply express the conditions on $\{F_n\}$ and $\{C_n\}$ stated in the corollaries of Section 4 in terms of similar conditions on $\{F_n\}$ and $\{P_n\}$.

In view of the widespread consideration of algorithms fitting the framework of (2.1.1), (1), and (2), it is of interest to reduce the moment conditions on $\{F_n\}$ and $\{P_n\}$ stated in the corollaries of Section 4 to moment conditions on $\{X_j\}$ and $\{s_j\}$. In case $\{X_j\}$ and $\{s_j\}$ are normally distributed, greatly simplified conditions can be established for this family of algorithms, as shown in Section (5.2).

(5.2) The normal case. Assume the same notation and structure as in Section 5.1. Further, assume that all random variables involved are jointly normally distributed. Then all of the "covariance decay rate conditons" for $\{F_n\}$ and $\{P_n\}$ stated in the corollaries of Section 4 and Section 5.1 can be expressed in terms involving the covariance functions of elements of $\{X_j\}$ and $\{s_j\}$. In order to accomplish this, some properties of joint moments of normally distributed random variables are required.

Let $z_1, z_2, \ldots, z_8$ be zero-mean jointly normally distributed with $E(z_i z_j) = \sigma(i,j)$. Using the properties of the characteristic function of $z_1, z_2, \ldots, z_8$, it is straightforward to show that

$$(1)\quad E\left(\prod_{i=1}^{4} (z_{2i-1} z_{2i} - \sigma(2i-1, 2i))\right)$$

$$= \sum_{\ell_1} \cdots \sum_{\ell_8} \sigma(\ell_1, \ell_2)\sigma(\ell_3, \ell_4)\cdots\sigma(\ell_7, \ell_8),$$

where the summation is over all possible ways of combining $\ell_1, \ell_2, \ldots, \ell_8 \in \{1, 2, \ldots, 8\}$, $\ell_i \neq \ell_j$ for $i \neq j$, into four distinct unordered pairs, no pair of which is (1,2), (3,4), (5,6), or (7,8). There are sixty terms in the sum.

With $F_n$ given by (5.1.1) and $K_n = K$, and defining $f(k,\ell) = \max_{1 \le i,j \le p} |(R_{xx}(k,\ell))_{i,j}|$, (1) can be applied to show that the conditions of Lemma (4.4) are satisfied if for some $\nu > 1/4$,

$$u^{\nu} \max_{1 \le i,j \le p} |(R_{xx}(k,k+u))_{i,j}|$$

is uniformly bounded for all non-negative integers k and u. Similarly, if for some $\nu_1 > 0$,

$$u^{\nu_1} \max_{1\le \ell \le p} \{|(P_s(k,k+u))_\ell|, |\rho_s(k,k+u)|\}$$

is uniformly bounded for all non-negative integers k and u, then the decay rate conditions stated in Corollary (4.5) on $\rho_F$, $\rho_C$, and $\rho_{FC}$ are all satisfied. Consequently, each member of the family of algorithms represented by (2.1.1) with $F_n$ and $P_n$ given by (5.1.1) and (5.1.2), and $K_n = K$ converge a.s. to $w_o$ provided that, in addition, $\{\mu_k\}$ and K are chosen as discussed in Section 5.1.

For example, let $\{x_k : k\epsilon\{0,\pm 1,\ldots\}\}$ be a real-valued zero-mean wide-sense stationary autoregressive-moving average (ARMA) normal random process. With $X_k = (x_k, x_{k-1}, \ldots, x_{k-p+1})'$ and $s_k = x_{k+\alpha}$ (integer $\alpha$), the required decay rates of (5.1.3) are easily established. Hence, each member of the family of algorithms represented by (2.1.1) with $F_n$ and $P_n$ given by (5.1.1) and (5.1.2), $K_n = K$, and $\{\mu_n\}$ satisfying Assumption (2.2) converge a.s. to $w_o$.

## 6. A simple almost sure convergence result.

(6.1) THEOREM. *Suppose that Assumption (2.1) holds and that there exist sequences $\{a_n\}_{n=1}^\infty$ and $\{b_n\}_{n=1}^\infty$ of non-negative real numbers (possibly random) satisfying $||F_n - R|| \le a_n$ and $|F_n w_o - P_n| \le b_n$. Furthermore, suppose that there exists a positive integer $n_o$ (possibly random) such that for all $n \ge n_o$, $0 < \mu_n \lambda_{min}(R) < 1$. Then for all $n \ge n_o$,*

$$(1)\quad |V_{n+1}| \le |V_{n_o}| \cdot \prod_{k=n_o}^{n} (1-\mu_k d_k) + \max_{n_o \le k \le n} (b_k/d_k) \cdot \Big(1-\prod_{j=n_o}^{n} (1-\mu_j d_j)\Big),$$

*where $d_k = \lambda_{min}(R) - a_k$. Furthermore, if $\sum_{k=1}^{n} \mu_k d_k \xrightarrow{a.s.} \infty$ and $b_n d_n^{-1} \xrightarrow{a.s.} 0$ as $n\to\infty$, then $|V_n| \xrightarrow{a.s.} 0$ as $n\to\infty$.*

PROOF. From Assumption (2.1),

$$V_{n+1} = V_n - \mu_n R V_n - \mu_n (F_n V_n + F_n w_o - P_n - R V_n),$$

so that for all $n \geq n_o$,

$$(2)\quad |V_{n+1}| \leq (1-\mu_n \lambda_{min}(R))|V_n| + \mu_n ||F_n - R|| \cdot |V_n| + \mu_n |F_n w_o - P_n|$$

$$\leq (1-\mu_n d_n)|V_n| + \mu_n b_n.$$

Iterating (2), for all $n \geq n_o$,

$$(3)\quad |V_{n+1}| \leq |V_{n_o}| \prod_{k=n_o}^{n} (1-\mu_k d_k) + \sum_{k=n_o}^{n} \prod_{j=k+1}^{n} (1-\mu_j d_j)\mu_k d_k (b_k d_k^{-1}).$$

Since all terms appearing in the sum in (3) are a.s. non-negative, (i) follows immediately from (3) with the aid of Lemma 1 of Albert and Gardner (1967, p. 189). Furthermore, if $\sum_{k=1}^{n} \mu_k d_k \xrightarrow{a.s.} \infty$ and $b_n d_n^{-1} \xrightarrow{a.s.} 0$ as $n \to \infty$, then (3) and the Toeplitz Lemma show that $|V_n| \xrightarrow{a.s.} 0$ as $n \to \infty$. □

(6.2) Remark. Theorem (6.1) is applicable to the case discussed in Section 5 with $K_k = k$.

ACKNOWLEDGMENT. The author wishes to express his sincere gratitude to Dr. L. L. Scharf for introducing the author to the problem, as well as for his continued direction throughout the course of this work. Further, the author is deeply indebted to a referee who discovered a major technical flaw in the original manuscript.

## REFERENCES


[1] Albert, A. E. and Gardner, L. A., Jr. (1967). *Stochastic Approximation and Nonlinear Regression.* M. I. T. Press, Cambridge, Massachusetts.

[2] Farden, D. C. (1975). Stochastic Approximation with Correlated Data. Ph.D. Dissertation, Department of Electrical Engineering, Colorado State Univ.

[3] Knopp, K. (1947). *Theory and Applications of Infinite Series.* Hafner, New York.

[4] Parzen, E. (1961). An Approach to Time Series Analysis. *Ann. Math. Statist. 32* 951-989.

[5] Robbins, H., and Monro, S. (1951). A Stochastic Approximation Method. *Ann. Math. Statist. 22* 400-407.

[6] Sakrison, D. J. (1966). Stochastic Approximation: A Recursive Method for Solving Regression Problems. In *Advances in Comm. Syst. 2* (A. V. Balakrishnan, ed.). 51-106. Academic Press, New York and London.

[7] Schmetterer, L. (1961). Stochastic Approximation. *Fourth Berkeley Symp. Math. Statist. Prob. 1* 587-609.

[8] Schmetterer, L. (1969). Multidimensional Stochastic Approximation. In *Multivariate Analysis II* (P. R. Krishnaiah, ed.) 443-460. Academic Press, New York and London.

[9] Serfling, R. J. (1970). Moment Inequalities for the Maximum Cumulative Sum. *Ann. Math. Statist. 41* 1227-1234.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle) (6) Stochastic Approximation with Correlated Data. | | 5. TYPE OF REPORT & PERIOD COVERED (9) Technical Rep[redacted]t. |
| | | 6. ~~PERFORMING ORG. REPORT NUMBER~~ |
| 7. AUTHOR(s) (10) David C. Farden Louis L. Scharf | | 8. ~~CONTRACT OR GRANT NUMBER(s)~~ (15) N66001-74-C-0035, N66001-75-C-0224 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Electrical Engineering, Colorado State University, Fort Collins, Colorado 80523 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (16) ARPA Order-1872 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency, 1400 Wilson Boulevard, Arlington, VA 22209 | | 12. REPORT DATE (11) Apr[redacted]976 |
| | | 13. NUMBER OF PAGES 22 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Naval Undersea Center, San Diego, California 92132 (12) 34p. | | 15. SECURITY CLASS. (of this report) Unclassified |
| | | 15a. DECLASSIFICATION DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Stochastic approximation, Robbins-Monro procedure, almost sure convergence, moment conditions, covariance decay-rate conditions, correlated data.

> W = W SUB O EPSILON $R_W^P$

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)


New almost sure convergence results are developed for a special form of the multidimensional Robbins-Monro (RM) stochastic approximation procedure. The special form treated can be viewed as a stochastic approximation to the solution $w = w_o \epsilon R^p$ of the linear equations $Rw = P$, where $R$ is a pxp positive definite symmetric matrix. This special form commonly arises in adaptive signal processing applications. Essentially, previous convergence results for the RM procedure contain a common "conditional expectation condition" which is


Next Page

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

406 434

20. con't.

extremely difficult (if not impossible) to satisfy when the "training data" is a correlated sequence. In contrast, the new convergence results incorporate moment conditions and covariance function decay rate conditions. The ease with which these results can be applied in many cases is illustrated.